

ACOUSTIC BEAMFORMING AND SPEECH RECOGNITION USING MICROPHONE ARRAY

PROJECT THESIS

**Under the guidance of
Prof. Lakshi Prosad Roy**

**Submitted By
Abhijeet Patra**

Arun Kumar Chaluvadhi





NATIONAL INSTITUTE OF TECHNOLOGY **DECLARATION**

We hereby declare that the project work entitled “**Acoustic Beamforming and Speech Recognition using Microphone Array**” is a record of our authenticated work under Prof. L.P. Roy, National Institute of Technology, Rourkela. Throughout this documentation wherever contributions of others are involved every endeavor was made to acknowledge this clearly with due reference to literature. This work is being submitted in partial fulfillment of the requirement for the degree of bachelor of engineering in electronics and communication/instrumentation engineering at National institute of technology, Rourkela for the academic session 2010-2014.

Abhijeet Patra
(110ec0178)

Chaluvadhi Arun Kumar
(110ei0586)



National Institute of Technology, Rourkela

CERTIFICATE

This is to certify that the thesis titled “**Acoustic Beamforming and Speech Recognition using Microphone Array**” submitted by **Abhijeet Patra** (110EC0178) and **Arun Kumar Chaluvadhi**(110EI0586). The work has been done for the partial fulfillment of the requirements of Bachelors in Technology in Electronics and Communication/Instrumentation, National Institute of Technology, Rourkela. This authentic work has been carried out under my supervision and guidance.

Prof. Lakhsi Prosad Roy,

Department of Electronics and communication.

Date:

ACKNOWLEDGEMENT

The project work has been completed successfully and we owe a number of people associated with the project.

First of all, we would like to thank Prof. Lakshi Prosad Roy for giving us the plausible opportunity to work under this compulsive topic.

We wish to extend our wholehearted thanks to Prof. S. K. Meher, Head of the department for approving our project.

We are also grateful to the Research scholars and M.Tech students for the helping hand.

Abstract

This report contains a piece of work on array signal processing for microphone array beamforming and its usability in NI PCI 4461 data acquisition system. Microphone arrays have great potential in practical applications of speech processing, due to their ability to provide both noise robustness and hands-free signal acquisition.

Here for sound and vibration analysis we require data acquisition systems and this data acquisition system consists of sensors DAQ systems and processor with programmable software and here we have used NI PCI 4461 system to study sound using two microphones.

Furthermore this report also presents the work on fundamental speech recognition process where we can verify that the speaker by testing phase and training phase.

Contents

DECLARATION	i
CERTIFICATE	ii
ACKNOWLEDGEMENT	iii
ABSTRACT	iv
CHAPTER 1: Introduction	1
1.1 Array Beamforming	2
1.2 Speech Recognition	2
CHAPTER 2: LITERATURE REVIEW	3
2.1 Array Processing Techniques	4
2.1.1 Sound and its propagation	4
2.1.2 Aperture or microphone	5
2.1.3 Linear sensor Array	6
2.1.3.1 Spatial Aliasing	7
2.2 Beamforming	8
2.3 Types of beamforming	10
2.3.1 Fixed Beamforming	10
2.3.1.1 Delay sum beamformer	10
2.3.2 Adaptive beamforming	12
2.3.2.1 Generalized side lobe canceller	12
2.3.3 DE reverberation Techniques	14
2.4 Data Acquisition Systems	14
2.4.1 Sound and vibration systems	15
2.5 Speech Recognition	16
2.5.1 Acquiring Speech	16
2.5.2 Analyzing the acquired speech	17
2.5.3 Developing a speech detection algorithm	17

CHAPTER 3: NI PCI 4461	18
3.1 Software requirements	19
3.2 calibration procedure	19
3.3 specifications of NI 4461	20
3.4 block diagram of NI 4461	21
3.4.1 Analog input	22
3.4.2 ADC	23
3.4.3 Analog output	24
CHAPTER 4: RESULTS	26
4.1 Effect of phase shift on angle of projection	27
4.2 effect of distance on beam width	28
4.3 effect of number of microphones on beam width	28
4.4 delay sum beamformer output	29
4.5 detecting mismatch between 2 speaker spectrums	29
CHAPTER 5: CONCLUSIONS	31
5.1 Conclusion	32
5.2 Future Scope	32
LIST OF FIGURES	33
REFERENCES	34

Chapter 1

INTRODUCTION

1. INTRODUCTION

1.1 Array beamforming

To get high SNR ratio of speech we bring microphones closer to the speaker which is not always possible, so in this situations we use array processing techniques. In this array processing techniques we use multiple microphones to get high SNR ratio. Here we use beamforming which is one of the array processing technique. In beamforming we are able to get speech from desired direction. Beamforming for speech application is different from radar and sonar applications in so many ways.

Generally in speech SNR is too small but positive and also ratio of direct and reverberation signal is negative. Speech is very wide band and changes its spectrum continuously which is dissimilar to radar and sonar application and quite often targets move rapidly or change their position which makes fixed beamforming unreliable. In those cases we use adaptive beamforming. In adaptive beamforming we decide look directions in such a way that, desired signal always have the ambient noise below a threshold level.

1.2 Speech recognition

Speech recognition is the process of recognizing who is speaking rather than what the speaker is speaking based on the information/data stored. This is done in two stages namely learning stage and the testing stage. In the training stage the speaker has to utter something to feed the data as speech samples. And in the testing phase the input is matched with the sample to validate the speaker. This process helps to authenticate or verify the identity of a person.

For speech recognition, we use data acquisition system which measures sound with computer and for sound and vibrations here we use NIPCI 4461 system.

Chapter 2

LITERATURE REVIEW

2.1 ARRAY PROCESSING TECHNIQUE

2.1.1 Sound and its propagation

Sound waves propagate through fluids as longitudinal waves. The molecules within the fluid pull away and forth within the direction of propagation, manufacturing regions of compression and growth. By victimization Newton's equations of motion to consider an infinitesimal volume of the fluid, an equation governing the wave's propagation is may be often developed. A generalized differential equation for acoustic waves is sort of advanced because it depends upon properties of the fluid, however, forward a perfect fluid with zero viscosity; the differential equation are often derived as

$$\nabla^2 x(t, r) - \frac{1}{c^2} \frac{\partial^2 x(t, r)}{\partial t^2} = 0$$

Where $x(t, r)$ is a function representing the sound pressure at a point in time and space and ∇^2 is the Laplacian operator. The speed of propagation c , depends upon the pressure and density of the fluid, and is approximately 330ms⁻¹ in air. The wave equation of Equation 1 is known as the governing equation for a wide range of propagating waves, including electromagnetic waves. The solution to the differential wave equation can be derived using the method of separation of variables. For sources at larger distances solution can be considered as plane waves and is given by

$$x(t, r) = -\frac{A}{4\pi r} e^{j(\omega t - kr)}$$

Where A is the wave amplitude, $\omega = 2\pi f$ is the frequency in radians per second, and the wavenumber vector k indicates the speed and direction of wave propagation.

2.1.2 Aperture or microphone

The term aperture is employed to refer to region that transmits or receives propagating waves. A transmitting aperture is said to be active aperture, whereas a receiving aperture is thought as a passive aperture. For example, in optics, an aperture may be a hole in an opaque screen, and in electromagnetics it may be an electromagnetic antenna. In acoustics, an aperture is an electroacoustic transducer that converts acoustic signals into electrical signals (microphone), or vice-versa (loudspeaker). There are two types of apertures.

1. Continuous aperture
2. Linear aperture

Consider a general receiving aperture of volume V where a signal $x(t, r)$ is received at time t and spatial location r . Treating the infinitesimal volume dV at r as a linear filter having impulse response $a(t, r)$, the received signal is given by the convolution as

$$x_R(t, r) = \int_{-\infty}^{\infty} x(\tau, r) a(t - \tau, r) d\tau$$

By taking Fourier transform

$$X_R(f, r) = X(f, r) A(f, r)$$

The term $A(f, r)$ is known as the aperture function or the sensitivity function, and it defines the response as a function of spatial position along the aperture. The response of a receiving aperture is inherently directional in nature, because the amount of signal seen by the aperture varies with the direction of arrival. The aperture response as a function of frequency and direction of arrival is known as the aperture directivity pattern or beam pattern. To simplify aperture directivity we take linear aperture of length L on x axis so spatial location r is simplified to x_a .

2.1.3 Linear sensor array

A sensor array can be considered to be a sampled version of a continuous aperture, where the aperture is only excited at a finite number of discrete points. The overall response of the array can be determined as the superposition of each individual sensor response for N elements is

$$A(f, \mathbf{x}_a) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} w_n(f) e_n(f, \mathbf{x}_a - \mathbf{x}_n)$$

Where $w_n(f)$ is the complex weight for element n, $e_n(f, \mathbf{x})$ is its complex frequency response or element function, and \mathbf{x}_n is its spatial position on the x-axis. If we apply directivity pattern to each individual microphone resultant directivity or array factor (AF) is given by

$$AF = \frac{\sin\left(\frac{N \times \psi}{2}\right)}{\sin\left(\frac{\psi}{2}\right)}$$

And

$$\psi = k \times d \times \cos \phi + \beta$$

Where

N=number of microphones

K=wave number

d= distance between microphones

ϕ =angle of arrival

β = initial phase shift between microphone

so angle of arrival is deduced from where maximum array factor will occur that is

$$\frac{\psi}{2} = \pm 2n\pi$$

By solving this we will get angle of arrival is

$$\phi = \cos^{-1} \frac{2n\pi - \beta}{kd}$$

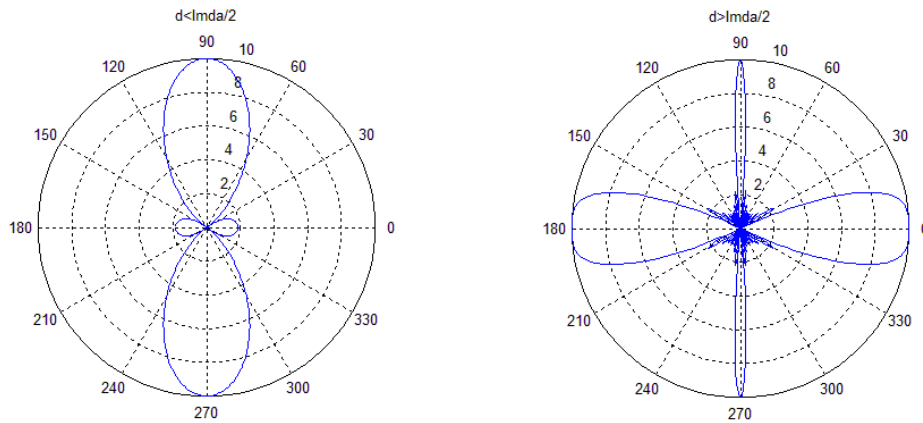
and beam width of the arrival is $2/L$ where $L=N \times d$ which effective width of microphone array and it is considered to be length of continuous aperture which it samples. By using an array of microphones rather than a single microphone, we are able to achieve spatial selectivity, reinforcing sources propagating from a particular direction, while attenuating sources propagating from other directions.

2.1.3.1 Spatial aliasing

Linear sensor array do spatial sampling. An analogous to Nyquist frequency, which is the minimum sampling frequency required to avoid aliasing in the sampled signal, a requirement exists to avoid grating lobes in the directivity pattern. If λ_{\min} is minimum wavelength of signal of interest then spacing between two adjacent microphones is given by

$$d < \lambda_{\min}/2$$

This is also known as spatial sampling and their directivity is given in following diagrams by changing distance between microphones

**FIGURE -1**

By observing this diagram it is clear that when distance between microphones is less than $\lambda/2$ there is more than one direction in which speech is arrived. These extra lobes are called grating lobes which are small in the case of d is greater than $\lambda/2$.

2.2 Beamforming

The most widely used array processing method is called beamforming. It refers to the any method that algorithmically steers the sensor in the array toward target signal. The direction is called look direction. Arbitrarily placed sensors can work as a microphone array and results in beamforming of acoustic signal.

Output from a microphone array is sum of all individual elements. In such case output from each individual microphone is delayed version of source signal and attenuated by a_i times (i represents i^{th} microphone) and also it contains some uncorrelated noise. so output from each microphone is

$$x_i(t) = a_i s_i(t - \tau_i) + v_i$$

By taking Fourier transform we get

$$X_i(f) = a_i S_i(f) e^{-j2\pi f \tau_i} + V_i(f)$$

If we consider all microphones then resultant output is

$$X(f) = d(f)S(f) + V(f)$$

Where

$$X(f) = [X_1(f) \dots X_i(f) \dots X_N(f)]^T$$
$$V(f) = [V_1(f) \dots V_i(f) \dots V_N(f)]^T$$

And $d(f)$ is steering vector is given by depends on distance between microphone and source

$$d(f) = \left[a_1 e^{j2\pi f \tau_1} \dots a_i e^{j2\pi f \tau_i} \dots a_N e^{j2\pi f \tau_N} \right]^T$$

To get desired signal output from all microphones is processed with frequency domain filters weights $W(f)$ and this is defined as

$$W(f) = [W_1(f) \dots W_i(f) \dots W_N(f)]^T$$

So resultant output is

$$Y = W^H X \text{ (H is hermitian transpose)}$$

For the output to be distortion less root mean square of noise should be minimum and also for desired look direction it should satisfy below equation. The well-known solution to this method is known as minimum variance distortion less response method (MVDR). Another popular method is least mean square (LMS)

$$W^H d = 1$$

2.3 Types of beamforming

There are different types of beamforming techniques are there. We choose any one of them according to our requirement and they are discussed below.

2.3.1 Fixed beamforming

Fixed beamforming is used when source is not moving and in this case weights of the filters are fixed so desired look direction is always same and examples of fixed beamforming are delay-sum-beamformer and filter sum beamformer which are discussed below.

2.3.1.1 Delay-sum beamformer

In delay-and-sum, signals from the various microphones are first time-aligned to adjust for the delays caused by path length differences between the target source and each of the microphones, using a variety of methods. The aligned signals are then summed together. Any interfering noise sources that do not lie along the look direction remain misaligned and are attenuated by the averaging. To align each channel in common time reference, a microphone location is commonly selected as a reference point. In frequency domain, delay signal is obtained by applying a phase shift to each channel signal spectrum. To normalize output after summation the signal channel is scaled by uniform gain factor that is number of microphones N and channel weight for delay sum is

$$W_i = \frac{1}{N} e^{-j2\pi f \tau_i}$$

Block diagram of delay sum beamformer is

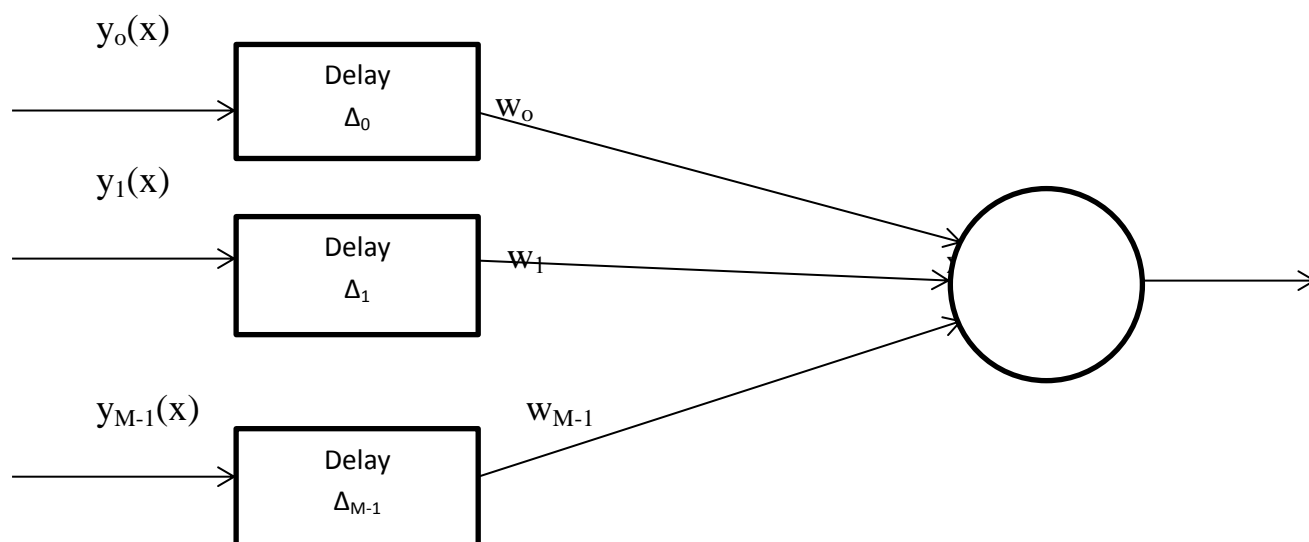


FIGURE-2(Delay-Sum beamformer)

Many microphone array-based speech recognition systems have successfully used delay-and-sum processing to improve recognition performance, and because of its simplicity, it remains the “method of choice” for many array-based systems. Most other array-processing procedures are variations of this basic delay-and-sum scheme or its natural extension, filter-and-sum processing, where each microphone channel has an associated filter and the captured signals are first filtered before being combined.

2.3.2 Adaptive beamforming

In adaptive beamforming, the array-processing parameters are dynamically changed according to some optimization criterion. In moving speaker environment where delay between two microphones changes we need to change weights according to surrounding changes or speaker position in those cases we use adaptive beamforming technique. The Frost algorithm is a weighted delay-and-sum technique in which the weights applied to each signal in the array are adaptively adjusted using constrained least mean square (LSM). Another method is the Griffiths-Jim algorithm or generalized side lobe canceller (GSC) which uses minimum variance distortion less response (MVDR). In some cases parameters are calibrated according to particular environment or user. In such cases noise reduction is updated.

2.3.2.1 Generalized side lobe canceller

GSC consists of two main structures. Upper part is fixed beam form which is non adaptive and lower part is adaptive beam form contains a blocking matrix and multiple input canceller. Blocking matrix realize the directional constraint by cancelling the signals from desired direction thus creating noise reference signals. Multiple input canceller uses unconstraint adaptive filters and further reduce noise from output of fixed beamformer in upper path. Block diagram of GSC is given as figure 3.

In real world applications we can't estimate microphone characteristics perfectly. In such cases desired signal may leak into blocking matrix results in attenuation of signal output from fixed beamformer because some of its output is deleted because of desire signal presence in blocking matrix so here we use adaptive blocking matrix. This type of generalized beamformer is known as robust generalized side lobe canceller (RGSC) and its block diagram is shown in figure 4

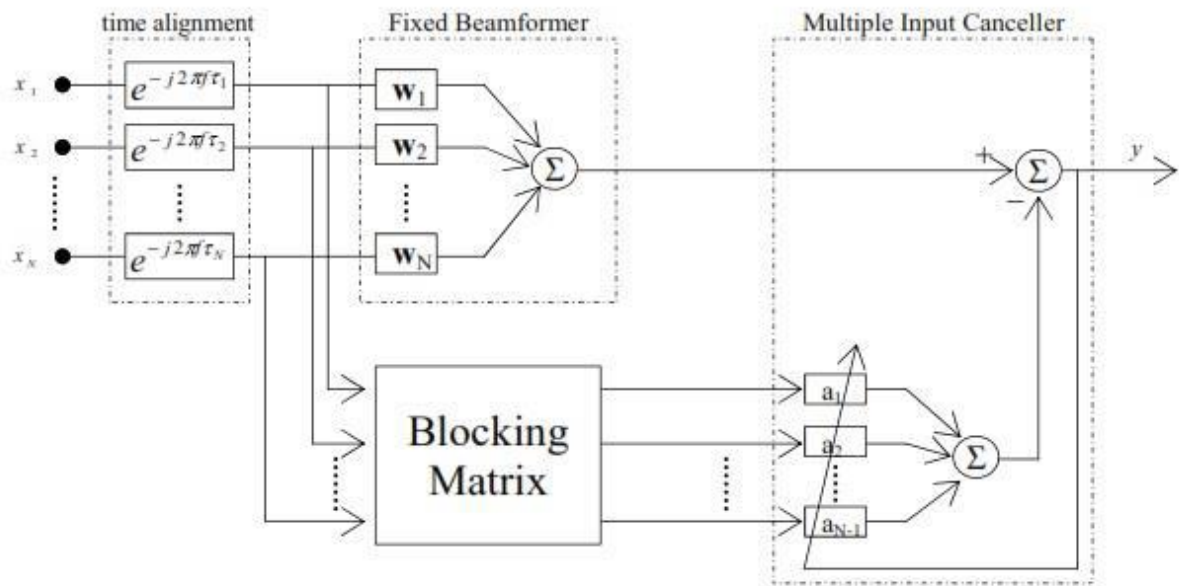


FIGURE -3(Generalized side lobe canceller)

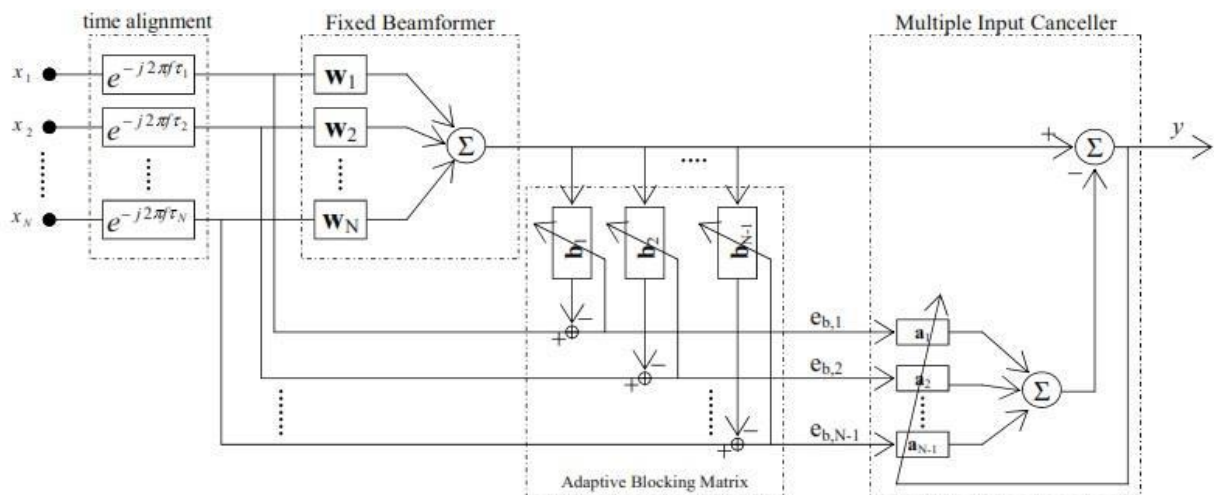


FIGURE -4(Robust Generalized side lobe canceller)

2.3.3 DE reverberation techniques

Reverberation is a main cause of poor speech recognition performance in microphone array-based speech recognition systems. Because none of the traditional beamforming methods successfully compensate for the negative effects of reverberation on the speech signal. So here we estimate room characteristics in which we use microphone. As room response results in non-minimum phase it is difficult to measure exact room response.

2.4 Data acquisition system

Data acquisition (DAQ) is the process of measuring an electrical or physical phenomenon such as voltage, current, temperature, pressure, or sound with a computer. A DAQ system consists of sensors, DAQ measurement hardware, and a computer with programmable software.

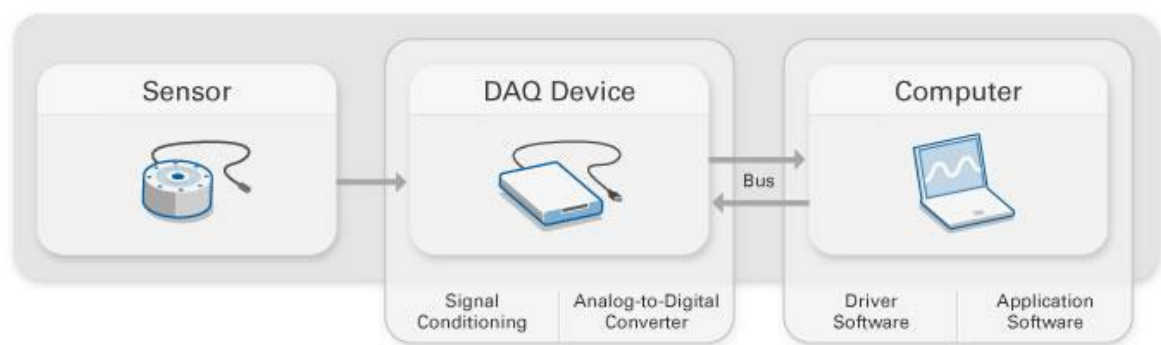


FIGURE-5(DAQ systems)

Compared to traditional measurement systems, PC-based DAQ systems exploit the processing power, productivity, display, and connectivity capabilities.

- **SENSOR:** The measurement of a physical phenomenon begins with a sensor. A sensor, also called a transducer, converts a physical

phenomenon into a measurable electrical signal. Depending on the type of sensor, its electrical output can be a voltage, current, resistance, or another electrical attribute that varies over time. Some sensors may require additional components and circuitry to properly produce a signal that can accurately and safely be read by a DAQ device.

- **DAQ SYSTEM:** DAQ hardware acts as the interface between a computer and signals from the outside world. It primarily functions as a device that digitizes incoming analog signals so that a computer can interpret them. The three key components of a DAQ device used for measuring a signal are the signal conditioning circuitry, analog-to-digital converter (ADC), and computer bus.
- **COMPUTER:** A computer with programmable software controls the operation of the DAQ device and is used for processing, visualizing, and storing measurement data

2.4.1 Sound and vibration system

- For measuring sound and vibration we use DAQ system.
- Required softwares are sound and vibrations software tool kit, sound and vibrations software measurement suite, sound and vibrations signal processing kit.
- It comes in two types of hardware packages.
 - i. PXI (high channel industrial platform)
 - ii. PCI (personal computer plug-ins)
- We can also use DAQ system it includes both software and hardware.

Here we use NI PCI 4461 DAQ system

2.5 Speech Recognition

Speech recognition is the process of recognizing who is speaking rather than what the speaker speaking is based on the information/data stored. This is done in two stages namely learning stage and the testing stage. In the training stage the speaker has to utter something to feed the data as speech samples. And in the testing phase the input is matched with the sample to validate the speaker. This process helps to authenticate or verify the identity of a person.

The workflow consists of three steps:

1. acquiring
2. analysis
3. interface development

2.5.1 Acquiring Speech

For training, speech is acquired from a microphone and brought into the development environment for analysis. For testing, speech is continuously streamed for processing. During the training stage, it is necessary to record repeated utterances of each digit in the dictionary. For example, we repeat the word 'one' many times with a pause between each utterance. Using the following MATLAB code with a standard PC sound card, we capture ten seconds of speech from a microphone input at 8000 samples per second.

We save the data to disk as 'mywavefile.wav'.

This approach works well for training data. In the testing stage, however, we need to continuously acquire and buffer speech samples, and at the same time, process the incoming speech frame by frame, or in continuous groups of samples.

The MATLAB code shown uses a Windows sound card to capture data at a sampling rate of 8000 Hz. Data is acquired and processed in frames of 80 samples. The process continues until the "RUNNING" flag is set to zero.

2.5.2 Analyzing the Acquired Speech

Starting with a word-detection algorithm that separates each word from ambient noise. Then deriving an acoustic model that gives a robust representation of each word at the training stage. Finally, we select an appropriate classification algorithm for the testing stage.

2.5.3 Developing a Speech-Detection Algorithm

The speech-detection algorithm is developed by processing the prerecorded speech frame by frame within a simple loop. For example, the MATLAB code continuously reads 160 sample frames from the data in 'speech'. To detect isolated digits, we use a combination of signal energy and zero-crossing counts for each speech frame. Signal energy works well for detecting voiced signals, while zero-crossing counts work well for detecting unvoiced signals. Calculating these metrics is simple using core MATLAB mathematical and logical operators. To avoid identifying ambient noise as speech, we assume that each isolated word will last at least 25 milliseconds.

.

Chapter 3

NI PCI 4461

3.1 Software Requirements

Before installing DAQ device physically we need to install proper device driver on calibrating computer. For that install DAQ mx 8.1 or later on calibrating computer. And this DAQ Assistant is compatible with version 8.2 or later of LABVIEW. First we need to install application software like matlab or labview and driver software like NI MAX. After installing these software, using lab view we can access the DAQ devices like NI PCI 4461.

3.2 Calibration procedure

- First confirm device recognition in NI max.
- Configure the device.
- Attach signal conditioning devices and sensors (TEDS).
- Run test panels
- Take DAQ measurement. NI 446X devices supports two types of calibration:

Self-calibration and external calibration.

- Self-calibration, also known as internal calibration, uses a software command and requires no external connections.
- Self-calibration improves measurement accuracy by compensating for variables such as temperature that might have changed since the last external calibration.
- NI 446x verification procedure verifies the accuracy prior to calibration.

- Based on above result perform external calibration.
- External calibration is generally performed with high-precision instruments at either NI or a metrology lab. This procedure replaces all calibration constants in the EEPROM and is equivalent to a factory calibration at NI.
- Perform another verification.
- Self-calibration retains the traceability of the external calibration

3.3 Specifications of NI 4461

- (a) Specifically designed for sound and vibration.
- (b) We need s&v suite installed on computer.
- (c) 24 bit resolution in ADC.
- (d) 118db dynamic range in DAC.
- (e) It can provide 6 gain settings (320mv to 42.4v).
- (f) 2 simultaneously sampled analog inputs at 204.8ks/s.
- (g) 2 simultaneously updated analog outputs at 204.8ks/s.
- (h) Variable anti-aliasing and anti-image filters.

3.4 Block diagram

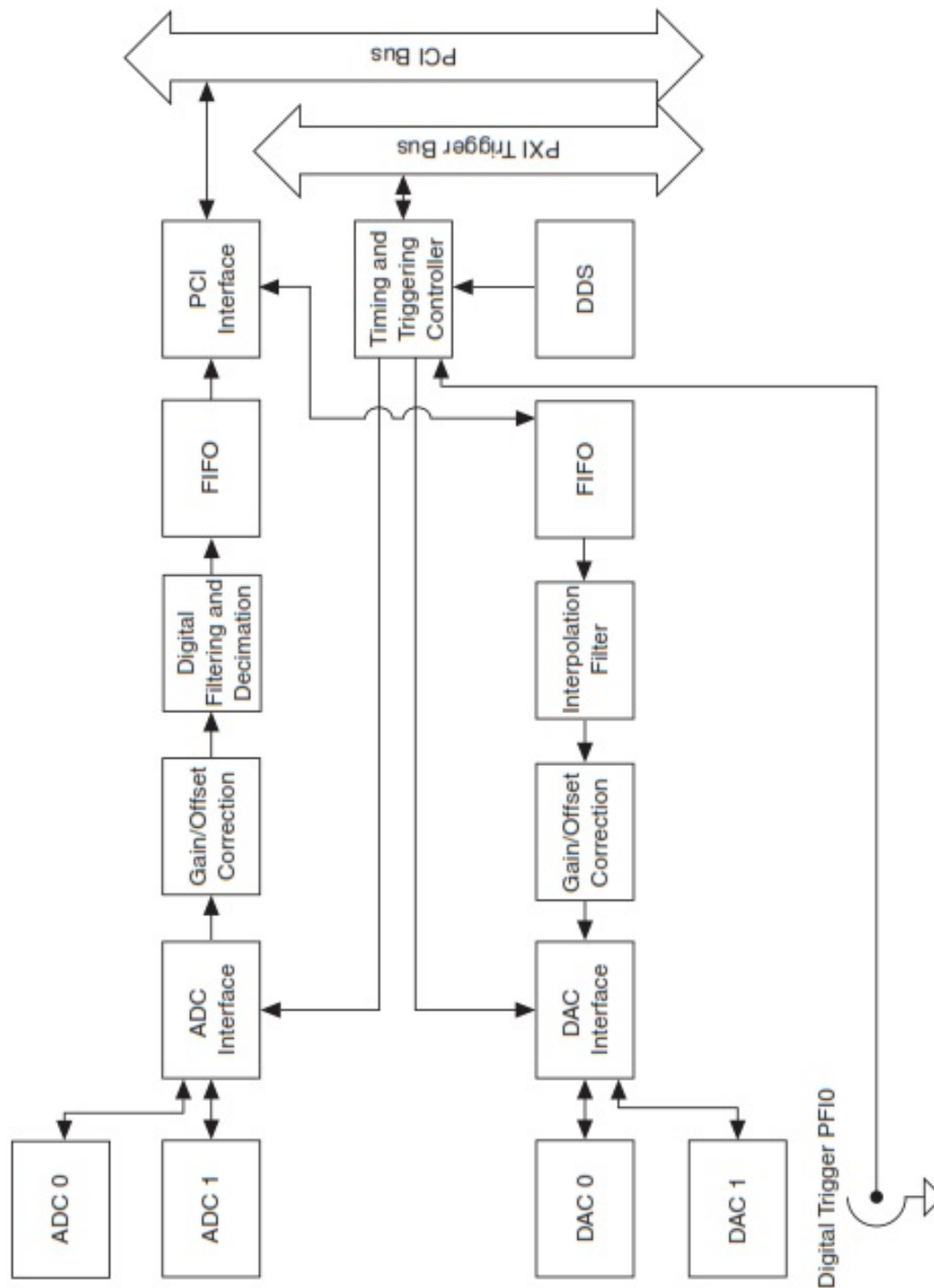


FIGURE-6(Block diagram of NI PCI 4461)

3.4.1 Analog input

The NI 446x supports two terminal configurations for analog input: differential and pseudo differential. The term pseudo differential refers to the 50 Ω or 1k Ω resistance between the outer connector shell and chassis ground. We need to choose channel configuration according to source input as given below

Source Reference	Channel configuration
Floating	Pseudodifferential
Grounded	Differential or pseudodifferential

The pseudo differential configuration provides a ground reference between the floating source and the DSA device by connecting either a 50 Ω or 1k Ω . For grounded source pseudo differential configuration is preferred (loop current by two ref grounds). You can configure them for AC/DC coupling. If you select DC coupling, any DC offset present in the source signal is passed to the ADC. The DC-coupling configuration is usually best if the signal source has only small amounts of offset voltage or if the DC content of the acquired signal is important. If the source has a significant amount of unwanted offset, select AC coupling to take full advantage of the input dynamic range. Selecting AC coupling enables a high pass resistor-capacitor (RC) filter into the signal conditioning path.

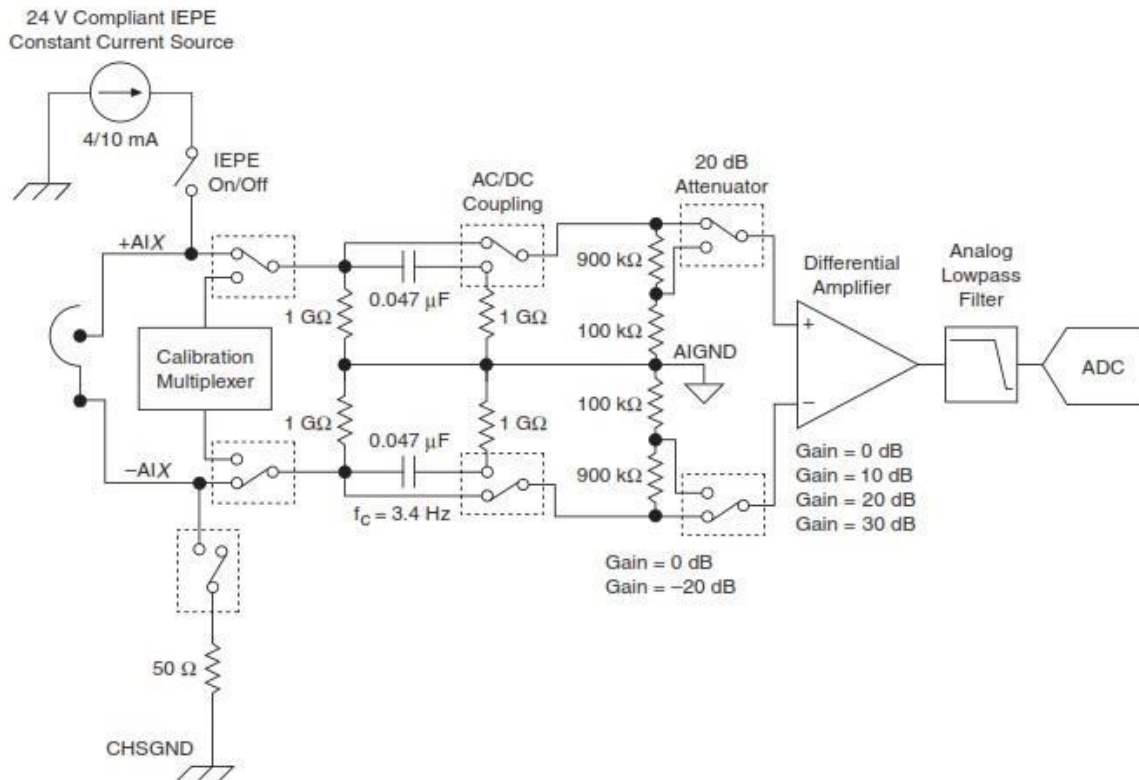


FIGURE-7(NI PCI 4461 analog input block diagram)

3.4.2 ADC

Each ADC in a DSA device uses a conversion method known as delta-sigma modulation. If the desired data rate is 51.2 kS/s, each ADC actually samples its input signal at 6.5536 MS/s, 128 times the data rate producing 1-bit samples that are sent to a digital filter. A digitizer or ADC might sample signals containing frequency components above the NY Quist limit. So we use anti-aliasing filters to eliminate components above the NY Quist frequency either before or during the digitization process can guarantee that the digitized data set is free of aliased components. The 1-bit, 6.5536 MS/s data stream from the ADC contains all of the information necessary to produce 24-bit samples at 51.2 kS/s.

- The delta-sigma ADC achieves this conversion from high speed to high resolution with a technique called noise shaping.
- The ADC adds random noise to the signal so that the resulting quantization noise, although large, is restricted to frequencies above the Nyquist frequency. This noise is not correlated with the input signal and is almost completely rejected by the digital filter.
- It uses a 1-bit DAC as an internal reference. As a result, the delta-sigma ADC is free from the differential nonlinearity (DNL).
- Over load detection is simple.

DSA device input channels share a FIFO buffer, and the output channels share a separate FIFO buffer.

3.4.3 Analog output

You can minimize output distortion by connecting the outputs to external devices with a high input impedance. Each output channel of the NI 4461 is rated to drive a minimal load of 600Ω .

- The NI 4461 supports two terminal configurations for analog output: differential and pseudo differential same as input.
- The delta-sigma DACs on the NI 4461 function in a way analogous to delta-sigma ADCs.
- The digital data first passes through a digital interpolation filter, then to the DAC resampling filter, and finally to the delta-sigma modulator.
- In the DAC, the delta-sigma modulator converts high-resolution digital data to high-rate, 1-bit digital data. As in the ADC, the modulator frequency shapes the quantization noise so that almost all of the quantization noise energy is above the Nyquist frequency.
- The digital 1-bit data is then passed to an inherently linear 1-bit DAC. The output of the DAC includes quantization noise at higher frequencies, and some images still remain near multiples of eight times the effective sample rate.

Analog output block diagram

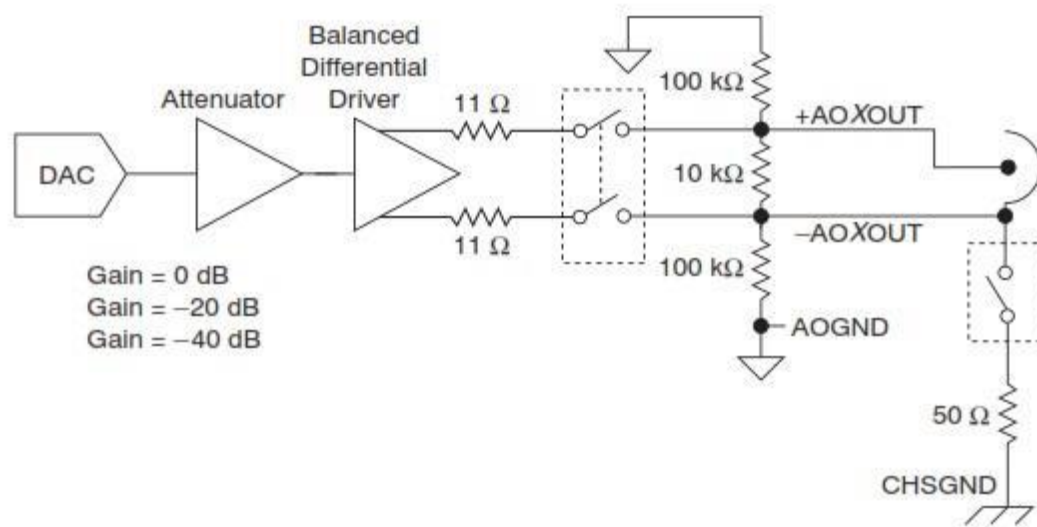


FIGURE-8(NI PCI 4461 analog output block diagram)

Chapter 4

Results

4.1 EFFECT OF PHASE SHIFT ON ANGLE OF PROJECTION as phase shift increases direction bends toward x axis.

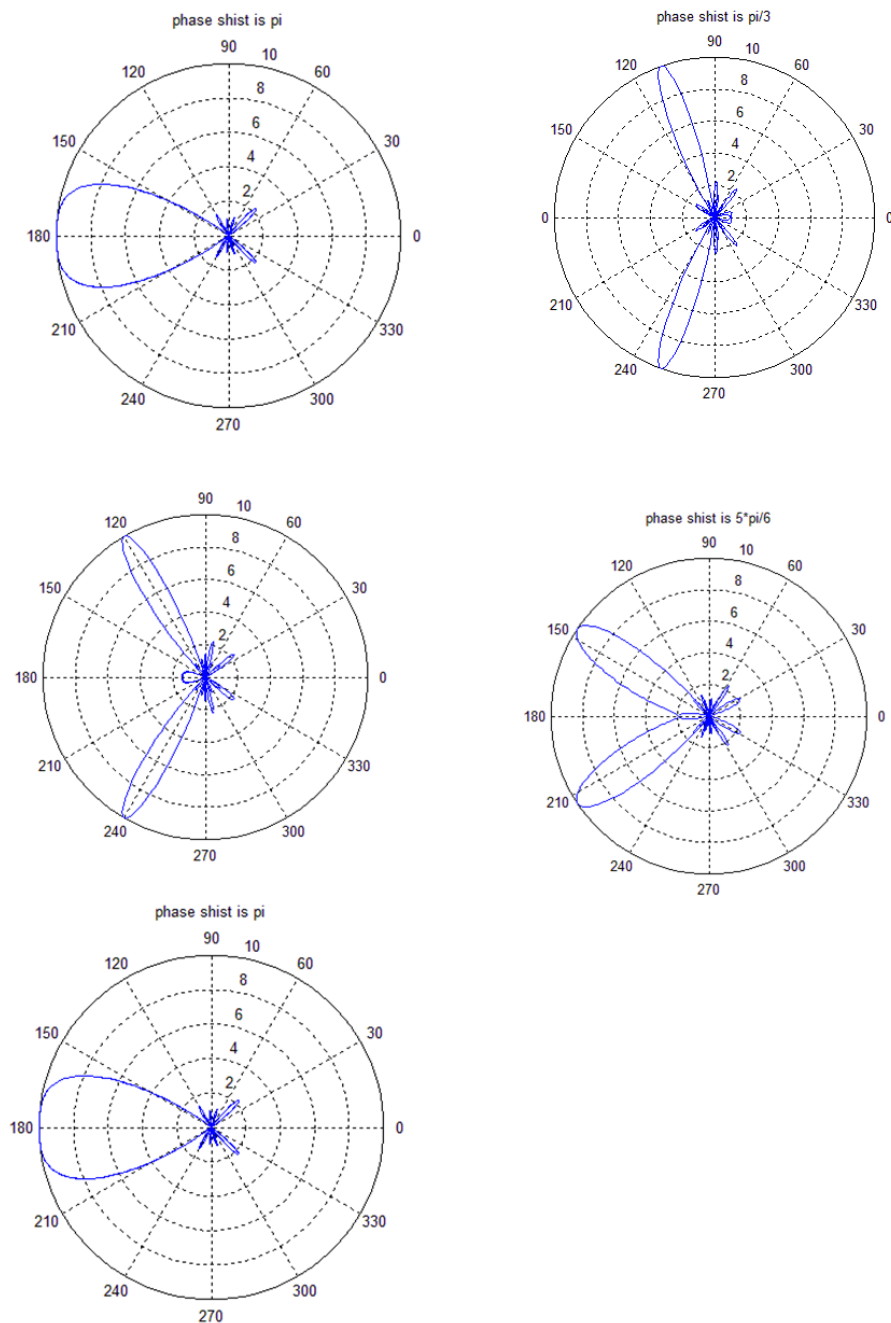


FIGURE-9(effect of phase shift on angle of projection)

4.2EFFECT OF DISTANCE ON BEAM WIDTH: As distance decreases beam width increases.

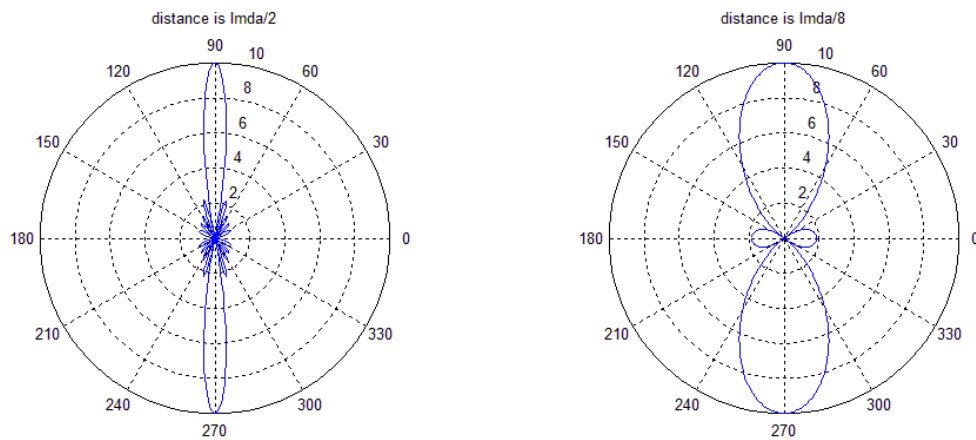


FIGURE-10(EFFECT OF DISTANCE ON BEAM WIDTH)

4.3EFFECT OF NUMBER OF MICROPHONES ON BEAMWIDTH: As microphones increases beam width decreases.

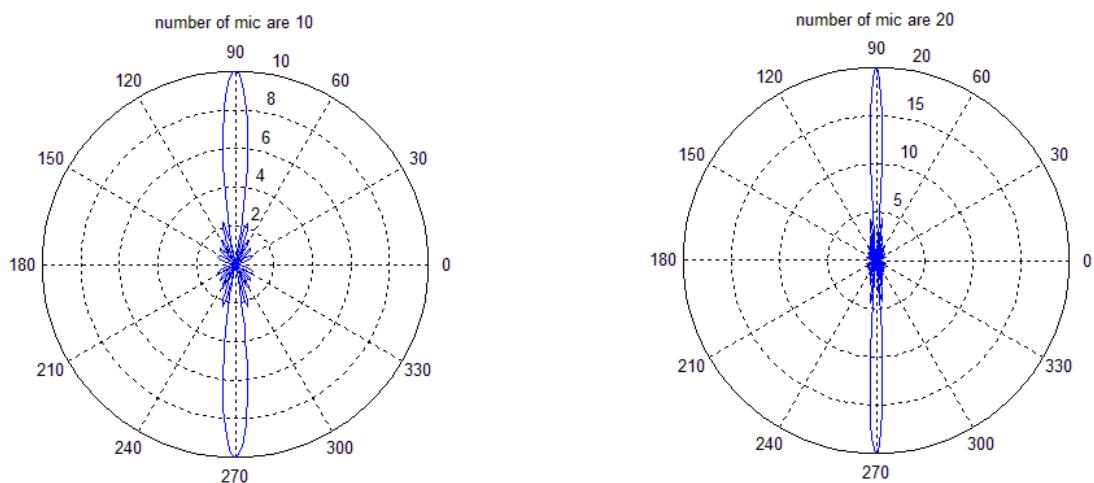


FIGURE-11(EFFECT OF NUMBER OF MICROPHONES ON BEAMWIDTH)

4.4 Delay sum beamformer effect:

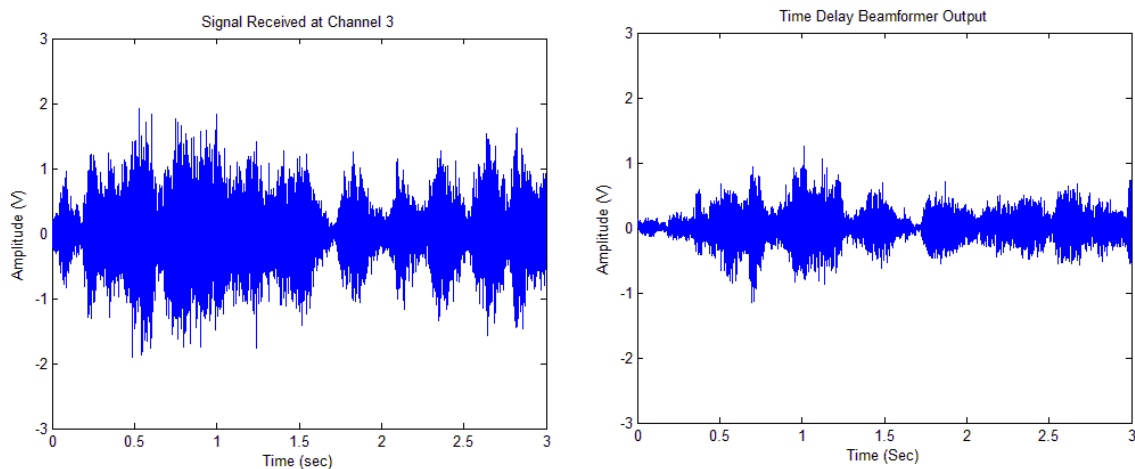
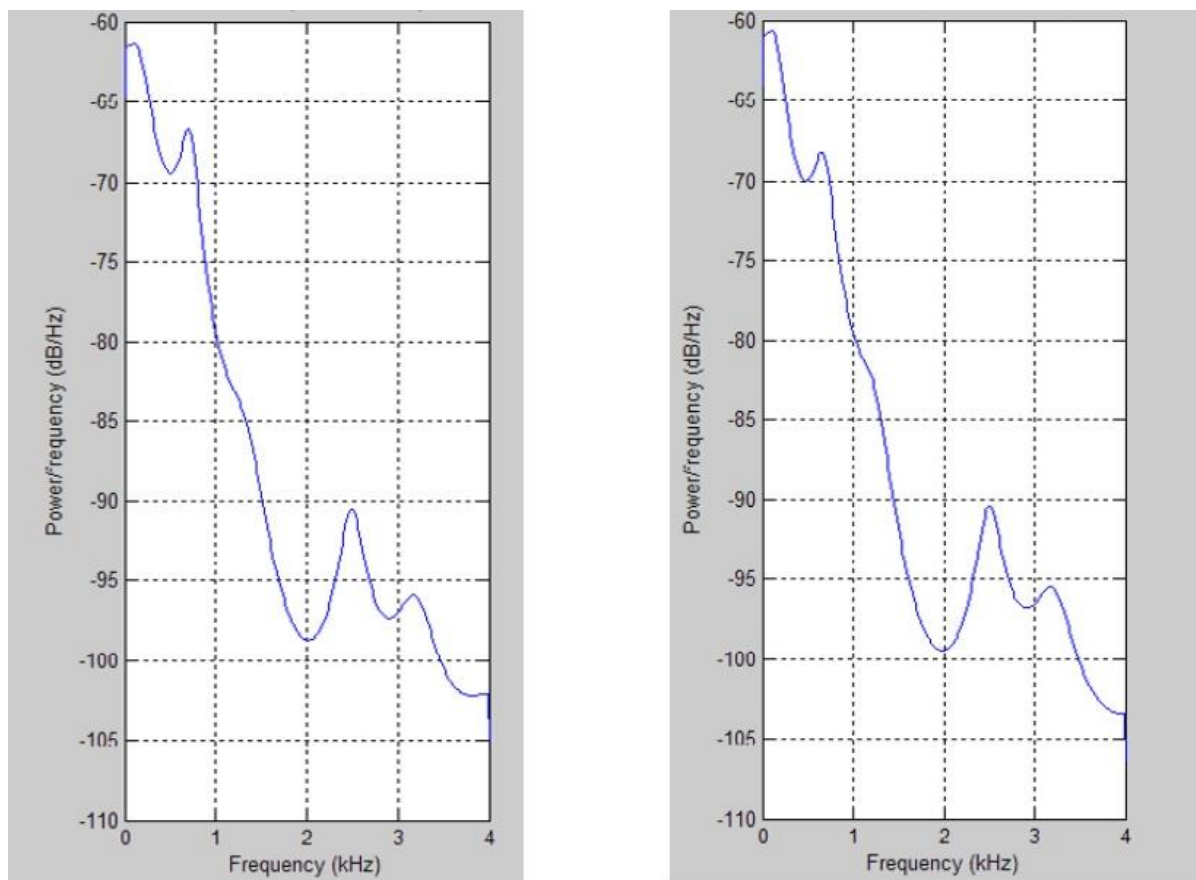


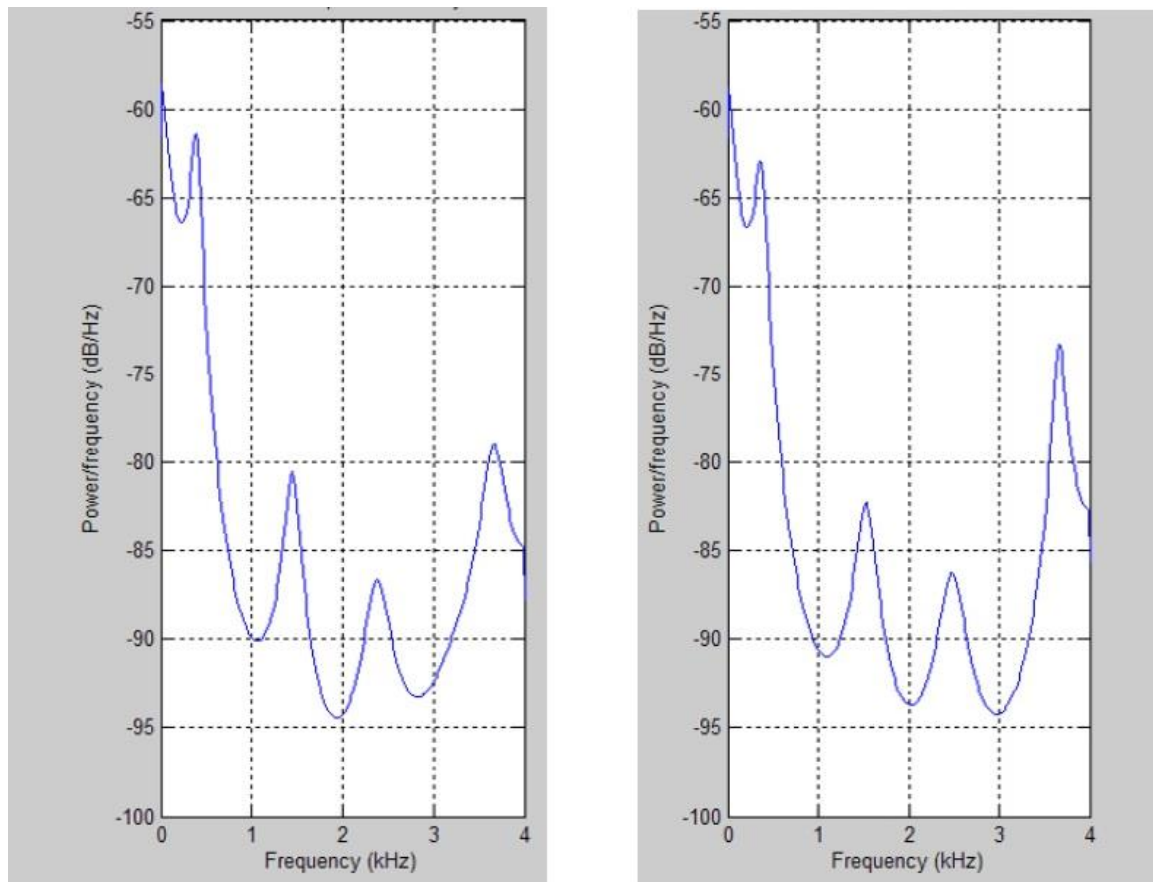
FIGURE-12(delay sum beamformer output)

4.5 Detecting mismatch between two speakers spectrum

Two different speakers uttering “one” (FIGURE-13a)



Two different speakers uttering “two”(FIGURE-13b)



CHAPTER 5

CONCLUSIONS

5.1 Conclusion

We showed the effect of distance between microphones on acoustic beamformer and also observed that beam width changes by changing number of microphones or distance between them. We successfully programmed delay-sum beamformer in matlab and by giving sample signals studied its effect on suppressing signal from other than look direction. And also spectrum mismatch between two speakers also studied to authenticate whether the person is the original speaker or not.

5.2 Future scope

- It can be used for detecting a particular speech in large auditoriums using speech mismatch.
- In similar fashion we can implement sending acoustic in particular direction.

LIST OF FIGURES

- Figure 1 Effect of distance on spatial aliasing
- Figure 2 Delay sum beamformer
- Figure 3 Generalized side lobe canceller
- Figure 4 Robust generalized side lobe canceller
- Figure 5 Data acquisition system
- Figure 6 Block diagram of NI PCI 4461
- Figure 7 NI PCI 4461 analog input block diagram
- Figure 8 NI PCI 4461 analog output block diagram
- Figure 9 Effect of phase shift on angle of projection
- Figure 10 Effect of distance on beam width
- Figure 11 Effect of number of microphones on beam width
- Figure 12 Delay-sum beamformer beam width
- Figure 13 Detecting mismatch between two speakers spectrums

Reference

- [1] <http://www.mathworks.in/help/phased> array system toolbox.
- [2] <http://www.ni.com/> data acquisition
- [3] Harvey F. Silverman, "Some Analysis of Microphone Arrays for Speech Data Acquisition" IEEE transactions on acoustics, speech, and signal processing, vol. assp-35, no. 12, December 1987 1699
- [4] Van Compernoee, Dirk "Switching Adaptive filters for enhancing noisy and reverberant speech using microphone array" IEEE transactions on acoustics, speech, signal processing 833-836 vol 2, published on April 1990
- [5] Ivan Himawan "Speech recognition using ad-hoc microphone arrays" published on April 2010